THE INFLUENCE OF ORTHOGRAPHY ON AUDITORY SPEECH PERCEPTION IN AMERICAN ENGLISH: EVIDENCE FOR INDIVIDUAL DIFFERENCES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

By

Nadezda Lvovna Dich

August 2008

ABSTRACT

A number of studies done in the past decade revealed influence of orthographic information on auditory speech perception in literate adults and children. Even though these studies yield convergent results, researchers in different languages do not always use the same experimental paradigms to assess orthographic effects. One of the goals of the present study was to extend the findings of orthographic effects in speech perception to American English using the same paradigm that has been previously used in French and Portuguese (simple auditory lexical decision task manipulating orthographic consistency of stimuli), thus strengthening the cross-linguistic validity of both the methodology and the results and laying the foundation for direct cross-linguistic comparisons and other extensions involving American English speakers.

Developmental research has shown that the size of orthographic effects in speech perception in children is associated with their reading level, however the question whether the size of orthographic effects observed in adult fluent readers is uniform has not been addressed. The second goal of the study was to explore individual differences in the size of orthographic effects and the nature of these differences. The results were convergent with findings from French and Portuguese: words with rhymes that have multiple spellings produced longer latencies and more errors than words with only one possible spelling. However the analysis of data from individual subjects showed that even in a sample that is quite homogeneous in age and literacy level (college students), there may be significant individual difference in the size of orthographic influences on speech perception, associated with differences in spelling competence.

BIOGRAPHICAL SKETCH

Nadezda Dich graduated from the Department of General Linguistics of Saint-Petersburg State University, Russia, in June 2003 with a diploma cum laude. In December 2003, she entered a PhD program in Institute for Linguistic Research of Russian Academy of Sciences. In August 2006, she transferred into a PhD program at the Department of Human Development at Cornell University. Nadezda's research interests include language acquisition, literacy in first and second language and reading and writing disorders.

ACKNOWLEDGEMENTS

I would like to thank my special committee members: prof. Barbara Lust, prof. Abby Cohn and prof. Elise Temple for their guidance in my research, as well as prof. Jim Booth for his help with statistical analyses, Sarah Spencer for her help with data collection and Eric Evans for technical support.

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

The impact of orthographic information on auditory speech processing has been drawing increasing attention from researchers in the field of speech perception and literacy. A number of studies done in the past decade provide converging evidence that orthographic information is automatically activated in spoken word recognition in literate people, just like phonological information is activated in reading (Taft, Castles, Davis, Lazendic, Ngyen-Hoan, 2008). These findings come from research on different languages employing different experimental tasks, which corroborates the validity of these findings and suggests that they reflect a real psychological phenomenon rather than being an artifact of a particular experimental setup in a particular language.

This phenomenon elicits increasing interest from different disciplines by providing insights on issues long in the focus of researchers' attention. Thus, from the cognitive perspective, it speaks to the question of modularity in information processing by demonstrating cross-modal interaction of subsystems supporting language. (for discussion, see Damian & Bowers, 2003). From the developmental perspective, it shows how an acquisition of a new skill – reading – modifies mechanisms underlying an existing set of skill – speech perception.

The influence of orthography on spoken language has long been recognized. However, until recently, orthographic effects in oral tasks were thought to be strategic, or conscious, and confined to the activities requiring explicit manipulations of sounds of words, such as phoneme counting task or phoneme deletion task[1] (see e.g. Tyler & Burnham, 2006, for review and discussion of orthographic influences on metaphonological tasks). The first study that tried to eliminate possible orthographic strategies and to show that activation of orthographic information in speech perception

[1] Participants are required to pronounce words without the first phoneme, e.g. *wage – age*.

is automatic was an auditory lexical decision study by Ziegler and Ferrand (1998) in which participants were not required to consciously reflect upon phonological structure of words. The authors showed that when listeners have to make a decision whether something they hear is a real word or not (a simple lexical decision task), they take longer to recognize words that have rhymes with multiple possible spellings (inconsistent words) than words that can only be spelled one way (consistent words)[1]. Ziegler and Ferrand argued that these results are unlikely to be caused by strategic activation of orthography, since the use of such orthographic strategies would not give participants any advantage to succeed in the task.

The results of Ziegler and Ferrand (1998) have since been replicated a number of times with different word sets in different languages (Ziegler, Petrova, & Ferrand, in press; Perre & Ziegler, 2008; Pattamadilok, Morais, Ventura, & Kolinsky, 2007; Ventura, Morais, & Kolinsky, 2007; Ventura, Pattamadilok, & Kolinsky, 2004; Ziegler, Ferrand, & Montant, 2004). These multiple successful replications make it unlikely that the effects interpreted as orthographic in these experiments were due to some unknown confound variable that was associated with orthographic consistency of words and that the experiment did not control for, which was one of the criticisms of the paradigm raised in literature (e.g., Taft et al., 2008).

A recent ERP study by Perre & Ziegler (2008) addressed another criticism, according to which orthographic effects observed in a simple lexical decision task can be interpreted as activation of orthography at the late decision making stage in this specific task, rather than as on-line activation of sublexical orthography during speech perception (Ventura et al., 2004, Cutler, Treiman, Van Ooijen, 1998). Using the same experimental setup as Ziegler and Ferrand (1998), Perre and Ziegler showed that the

---

[1] This paradigm will be discussed in more detail in the section 2.

ERP differences between consistent and inconsistent words could be observed as early as 320 ms from word onset – long before the end of the word. This finding is interpreted as suggesting that orthography interferes with speech processing at an early sublexical as well as at a late decisional stage and that this interference is not specific to the lexical decision task.

Results suggesting automatic activation of orthography in speech perception also come from experiments with different designs. Thus, lexical decisions have been shown to be faster when spoken words have a dense orthographic neighborhood[1] than when the orthographic neighborhood is sparse (Ziegler & Muneaux, 2007; Ziegler, J.C., Muneaux, M., & Grainger, 2003). Similarly, primed lexical decisions are faster if the target and prime overlap in both phonology and orthography than when the overlap is in phonology only (Taft et al., 2008; Chereau, Gaskell, & Dumay, 2007; Millier & Swick, 2003; Slowiaczek, Soltano, Wieting, & Bishop, 2003).

Although automatic activation of orthography in auditory speech perception has been demonstrated in a number of studies, extremely little is still known about the nature of this phenomenon, the course of its development in beginner readers and its manifestation in languages with different orthographic systems. Questions regarding these issues are only starting to be explored in the literature. For example, Pattamadilok et al. (2007) used a simple lexical decision paradigm manipulating orthographic consistency of stimuli to show that orthographic effects in speech perception can be observed cross-linguistically, but that their size depends on the overall transparency of the orthographic system[2], so that in a language with a less transparent orthography (French), orthographic effects in speech perception are

[1] The orthographic neighborhood of a word is the number of words spelled similarly to that word. See also section 3.1.1. for more detail.
[2] Transparency of orthographic system is the degree of consistency of mappings between sounds and orthographic units.

stronger than in a more orthographically transparent language (Portuguese). The authors interpreted these results as suggesting that the higher overall inconsistency of sound-to-spelling mappings in French makes French listeners more sensitive to the orthographic inconsistency of stimuli in auditory speech perception.

In order to make more cross-linguistic comparisons possible, more data are needed from other languages obtained using the same experimental design. At the onset of the present study, French and Portuguese were the only two languages in which a simple lexical decision task was used to study orthographic effects in spoken word recognition. In addition to the present study, a simple lexical decision study has since been done in British English and is only about to be published (Ziegler et al., in press, Experiment 3). Thus, one of the goals of the present study was to obtain converging evidence on orthographic effects in simple auditory lexical decision from American English. Extending the evidence to yet another language would add to cross-linguistic validity of both the methodology and previously obtained results and would lay a foundation for cross-linguistic comparisons of the sizes of orthographic effect in speech perception between American English and other languages.

There is at least one reason to believe that American English speakers may differ from both French and Portuguese speakers in the size of orthographic effects in speech perception. English differs in orthographic transparency from both French and Portuguese: while being less consistent than Portuguese but slightly more consistent than French in phonology-to-orthography mappings (Ziegler; Stone, & Jakobs, 1997), English is much more inconsistent in the opposite direction, i.e., in orthography to phonology mappings. This means that printed English words are more likely to have multiple possible readings than French or Portuguese words are (Ventura et al. 2004; Ziegler et al., 1997; Ziegler & Goswami, 2006). For example, *tear* can be read as /tir/ and /tɛr/, *wind* can be read as /wɪnd/ and /waɪnd/.

In addition to comparing the size of orthographic effects in speech perception in different languages, researchers are also starting to explore the nature of this phenomenon from a developmental perspective. Two recent studies compared beginner readers and advanced readers in order to determine at which point in the course of children's literacy acquisition orthography will start to influence auditory speech perception (Ziegler & Muneaux, 2007; Ventura et al., 2007). These studies obtained convergent results showing that whereas the pre-reader control group (Ventura et al., 2007), poor 1^{st} grade readers and dyslexic children whose reading level matched that of first graders' (Ziegler & Muneaux, 2007) did not exhibit any influence of orthography on auditory lexical decisions, orthographic effects were observed as early as 1^{st} grade for the best readers in that age group (Ziegler & Muneaux, 2007), and were reliably observed starting 2^{nd} grade (Ventura et al., 2007). These results suggest that orthographic information starts influencing spoken language processing almost as soon as orthographic representations are formed in young readers' memories.

Thus, research has shown that orthography does not affect children's speech perception in a uniform manner. Rather, its effects are determined by the child's reading level. In studies with adult fluent readers to date, however, participants have been considered as a homogeneous group for which the size of orthographic effects was calculated by averaging results over all subjects. None of the previous studies has addressed the question whether there may be differences among adults in how strong the involvement of orthography in speech perception is. This question is not ungrounded, though. Considerable differences in reading and spelling skills exist not only in child, but also in adult populations (e.g., Shankweiler, Lundquist, Dreyer, & Dickinson, 1996). Therefore one might expect to find individual differences in the size of orthographic effects on speech processing, associated with reading and spelling competence. This hypothesis is consistent with the finding that adults with a history of child dyslexia, even those who have overcome the reading deficit and score within norms on reading and spelling tests, exhibit less orthographic interference in metalinguistic tasks, such as phoneme counting (Bruck, 1992).

The second goal of this study was, therefore, by replicating orthographic effects in simple lexical decision task in American English normal readers, to lay the foundation for the comparison of orthographic effect sizes in normal readers and people with reading disorders. English is especially well-suited for studying people with reading disorders: because of its high inconsistency of spelling-to-sound mappings, English is harder to learn to read than other alphabetic languages (Ziegler & Goswami, 2006; Frith, Wimmer, & Landerl, 1998), as seen in English speaking children's poorer reading skills compared to their age-matched peers speaking other languages (Seymour, Aro, & Erskine, 2003) and in higher rates of dyslexia in English speaking countries (Lindgren, De Renzi, & Richman, 1985).

Finally, the third goal of the study was to see if individual differences in the size of orthographic effects in speech perception can be observed in adult normal readers, and if so, whether these differences would be correlated with individuals' spelling proficiency. Since it has been proposed that the presence of orthographic effects in child listeners is conditioned by the strength of their orthographic representations (Ziegler & Muneaux, 2007), it seems reasonable to hypothesize that good adult spellers may experience stronger orthographic effects in speech perception than poor spellers.

In order to meet all these goals, the study included two components:

- A simple lexical decision experiment manipulating orthographic consistency of stimuli rhymes.
- A spelling test that was aimed at assessing the memory for irregular lexical spellings and the knowledge of sublexical orthography-phonology correspondences –

the two aspects of spelling proficiency typically assessed in spelling research (see, e.g. Perry & Ziegler, 2004, Kreiner & Gough, 1990, for the discussion of lexical and sublexical processing in spelling).

The next section will describe the auditory lexical decision paradigm manipulating orthographic consistency of stimuli rhymes and present a comparative analysis of all results coming from different studies using this paradigm. The specific details of the materials and apparatus used in the present study as well as the results of the present study will be described in sections 3 and 4. Finally, section 5 will discuss the results of the present study in relation to previous findings.

2. METHODOLOGICAL BACKGROUND

In an auditory lexical decision paradigm manipulating orthographic consistency of stimuli rhymes, participants are presented with spoken words and nonwords in random order and are instructed to press 'yes' and 'no' accordingly as fast as possible. Half of the words and half of the nonwords are *consistent* and the other half are *inconsistent*. Consistent items are items that have a rhyme with only one possible spelling (e.g., *globe* is a consistent word because /ob/ is always spelled as -*obe* in English). Inconsistent items have a rhyme with multiple possible spellings (e.g., *name* since /em/ can be spelled either as -*ame,* as in *flame*, or as -*aim*, as in *claim*). Consistent and inconsistent items are matched on a number of variables potentially affecting reaction time, such as frequency, length, orthographic and phonological neighborhood density and uniqueness point (see section 3.1.1. for explanation of these variables).

Accuracy and reaction time (RT) of responses are measured for consistent and inconsistent words and nonwords. The statistical analysis is performed on correct mean RTs and error rates with subjects and items as random variables.

Table 1. Reaction times (standard errors) and error rates for consistent and inconsistent items, difference between consistent and inconsistent items and its significance in by-subject and by-item analyses in auditory lexical decision studies manipulating orthographic consistency of stimuli rhymes.

Source	Language	Con Words	Inc Words	Difference	Con Nonw	Inc Nonw	Difference	
Ziegler & Ferrand (1998)	French	756 (10.7)	818 (11.4)	62 ms $p < .001$ by subj $p < .005$ by item	859 (14.3)	864 (15.1)	5 ms n. s.	REACTION TIMES
Ziegler et al. (2004)	French	741 (7.8)	811 (7.6)	70 ms $p < .001$ by subj $p < .001$ by item	N/A	N/A	N/A	
Pattamadilok et al. (2007)	French	867 (1.9)	928 (2.6)	61 ms $p < .0001$ by subj $p < .001$ by item	1100 (4.7)	1135 (5.7)	35 $p = .001$ by subj $p = .05$ by item	
Ventura et al. (2004)[1]	Portuguese	882 (4.6)	934 (4.7)	52 ms $p < .001$ by subj $p < .05$ by item	1090	1058	-32 n. s.	
Ventura et al. (2007)	Portuguese	884 (3.9)	922 (3.6)	38 $p < .005$ by subj $p < .05$ by item	1017	1015	n. s.	
Ziegler et al. (in press)	British English	894 (n/a)	939 (n/a)	45 ms $p < .0001$ by subj $p < .05$ by item	N/A	N/A	N/A	
Source	**Language**	**Con Words**	**Inc Words**	**Difference**	**Con Nonw**	**Inc Nonw**	**Difference**	
Ziegler & Ferrand (1998)	French	7.8%	20.9%	13.1% $p < .001$ by subj $p < .05$ by item	11.0	13.1	2.1% n.s	ERROR RATES
Ziegler et al. (2004)	French	12.8%	22.3%	9.5% $p < .001$ by subj $p < .001$ by item	N/A	N/A	N/A	
Pattamadilok et al. (2007)	French	2.6%	3.6%	n. s.	3.7%	3.9%	n. s.	
Ventura et al. (2004)	Portuguese	9.5%	19.1%	9.6% $p < .0005$ by subj $p = .07$ by item	11.5	11.2	n. s.	
Ventura et al. (2007)	Portuguese	2.3%	11%	8.7% $p < .0001$ by subj	10.3%	11.5%	n. s.	
Ziegler et al. (in press)	British English	10.9%	10.2%	n. s.	N/A	N/A	N/A	

Table 1 presents the results of the original Ziegler & Ferrand (1998) study, its published replications in French and Portuguese and its unpublished replication in British English (Ziegler et al., in press).

To summarize, in all of the reviewed studies, participants tended to take on average longer to recognize inconsistently spelled words than consistently spelled words. The effect of orthographic consistency on reaction times was reliably observed for words in all studies and was significant in both by-subject and by-item analyses. Effect of consistency on error rates for words was observed less reliably. Only one study reported an effect of consistency on RTs for nonwords and no studies reported an effect of consistency on error rates for nonwords.

3. STUDY 1

3.1. METHOD

3.1.1. STIMULI AND DESIGN

3.1.1.1. Lexical Decision Task

Variables Manipulated and Controlled

A list of 120 stimuli was created for the lexical decision task. The list consisted of 60 monosyllabic English words and 60 monosyllabic nonwords. Half of the items were orthographically consistent and the other half were inconsistent. Information on orthographic consistency of rhymes was obtained from Ziegler, Stone & Jacobs's (1997) database of English monosyllables.

Words. Consistent and inconsistent words were matched on a number of characteristics that have been previously shown to affect reaction times of lexical decisions and were therefore controlled for in auditory lexical decision studies reviewed in section 2: frequency, length (in phonemes and letters), acoustic duration, orthographic and phonological neighborhood density, uniqueness point, onset phonotactic probability, onset orthographic and phonological consistency (Table 2).

Frequency. The information on frequency was taken from multiple sources: Kučera and Francis (1967), Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996), and CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). Multiple sources were used in order to obtain more objective estimates on frequencies than any single one could provide: the Kučera and Francis (1967) database that is traditionally used in psycholinguistic research is rather old and it is based on a small corpus (1,014,000 words), which makes it not very reliable (see, e.g., Zevin & Seidenberg, 2002, for discussion). HAL is based on a large corpus (131 mln. words) and its norms are more up-to-date (data were collected in 1995), but it is limited in types of written texts it is based on – it was all gathered from Usenet newsgroups (Lund & Burgess, 1996). Finally, CELEX is based on a wide variety of texts, however most of them are British English – only 15% are American English.

Length and duration. The two groups of words were matched on the number of letters and phonemes. The phoneme count was based on phonemic transcriptions provided in the English Lexicon Project database (Balota et al., 2002) available online at http://elexicon.wustl.edu/default.asp. The acoustic duration of each word was also measured after the stimuli were recorded. The two groups did not differ significantly in auditory duration.

Orthographic and phonological neighborhood density. Neighborhood density of a word is the number or words that can be obtained by changing one letter (orthographic neighbors) or one phoneme (phonological neighborhood) of the word. For example, *cat* and *eat* are orthographic neighbors, *beer* and *pier* are phonological neighbors. The information about neighborhood density was obtained from the English Lexicon Project database (Balota et al., 2002).

Uniqueness Point. Uniqueness point (UP) is the number of phonemes in the initial portion of the word that makes the word uniquely identifiable (i.e., no other word starts with the same phoneme sequence). Uniqueness point was calculated based on phonemic transcriptions provided in CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). Differences between British and American pronunciation were taken into account.

Phonotactic Characteristics. Stimuli were matched on the positional frequency of the initial phone (i.e., how many other words start with the same phone) and onset transitional probability, i.e., co-occurence probability of first two segments. This information was obtained from online Phonotactic Probability Calculator (Vitevitch, & Luce, 2004).

Onset Consistency. Orthographic and phonological consistency of onsets was calculated as type frequency of a particular spelling/pronunciation of an onset divided by the type frequency of all spellings/pronunciations of the onset. The calculations were based on CELEX database with necessary adjustments to American English pronunciation.

Table 2. Stimuli characteristics, Study 1

	CONSISTENT	INCONSISTENT	SIGNIF PROB
WORDS			
# friends[1]	2.4	2.3	.83
freq of friends	60.0	60.2	.99
# enemies[2]	0	12.6	
freq of enemies	0	1483.1	
type consistency ratio[2]	1	.06	
token consistency ratio[3]	1	.14	
freq Kučera and Francis	13.3	13.6	.93
freq HAL	4716.1	4771.3	.99
# of letters	5.0	4.6	.17
#of phonemes	3.9	3.8	.48
duration (ms)	570	593	.36
# of orth neighb	4.8	3.3	.13
# of phon neighb	10.8	14.8	.09
# init phone freq	0.07	0.07	.38
# init biphone freq	0.006	0.006	1.00
uniqueness point	4.0	4.1	.84
onset orth consistency	.94	.90	.52
onset phon consistency	.96	.97	.53
NONWORDS			
duration (ms)	543	553	.64
rhyme frequency (type)	208.6	211.3	.95
rhyme frequency (token)	8.0	7.1	.56

Nonwords. A list of nonwords was created by randomly exchanging the onset and rhymes of real monosyllabic words. Only rhymes not used for the word stimuli were used for nonwords. Half of nonwords had consistent rimes, whereas the other

[1] A *friend* is a word that has the same rime spelled the same way; an *enemy* is a word that has the same rime spelled differently (e.g., *line* and *spine* are friends, *line* and *sign* are enemies).
[2] Consistency ratio (token and type) is value ranging from 0 to 1 calculated as the number (or summed frequency) of *friends* of the word divided by the summed number (or summed frequency) of its *friends* and *enemies*.

half had inconsistent rimes. The two groups of nonwords were matched on the ***type and token frequency of the rhyme***. ***Duration*** of nonwords was measured after recording and editing the stimuli. The two groups of words did not differ significantly in duration (Table 2).

The complete list of stimuli is given in Appendix A.

Recording and Programming

Stimuli were recorded in a sound proof room by a female speaker from California, who read the words and nonwords in a sentence frame 'say ___ again'. The sentence frame was used in order to keep intonation constant and to control for speed. The stimuli were then excised from the sentences using *Praat 4.6.21*. The characteristics of recording equipment are listed in Appendix B.

The presentation of stimuli and data collection were controlled by *Presentation 12.1.* run on an Intel Celeron PC computer (see Appendix B for more information on the hardware). The software was programmed to present stimuli with a 2 second interval and record reaction times of responses (from the onset of auditory stimulus presentation) and accuracy of responses.

3.1.1.2. Spelling Test

The test consisted of two components intended to assess two different aspects of spelling competence: word-specific memory and the knowledge of sublexical regular orthography-to-phonology correspondences (Kreiner & Gough, 1990).

The first part consisted of 30 low-frequent (mean frequency ~ 2 per million) words with highly irregular spellings. Two sources were used to create the list. Monosyllables were taken from the Ziegler et al.'s (1997) database. Only those words that had 0 friends were included, which means that no other words with the same phonological rhyme have it spelled the same way in English (e.g., *plaid, suede, ache*). Polysyllables were taken from an online resource for teaching phonics to adult ESL (English as second language) learners (http://literacy.kent.edu/Midwest/Materials/ndakota/soup/index.html). This database lists the most regular spellings for each phoneme in English and gives lists of exceptions and rare spellings (e.g., *syrup*, *coyote*, *khaki*).

The second part of the test consisted of 40 nonwords. Because nonwords do not have a lexical spelling, the spellers can only rely on existing regularities in sublexical orthography-to-phonology correspondences in English, thus demonstrating their knowledge of those correspondences. The list of nonwords was created by combining existing English onsets and rhymes, so that the nonwords would be phonotactically legal.

A female speaker from upstate New York recorded the test items. The recording was edited in *Praat 4.6.21* to remove extraneous noises and create a single sound file with the words followed by the nonwords and each item repeated twice with a 2 seconds interval between repetitions and between items.

3.1.2. PROCEDURE

Participants were tested individually. Before the onset of the study participants read and signed the consent from and filled out a short survey containing questions pertaining to languages spoken and read and their acquisition. More specifically, the survey asked if participants spoke or read languages other than English and if so, at what age those languages were acquired and how proficient the participants were in those languages. The survey also asked participants whether they remembered by what method they were taught reading English in school (phonics vs. whole word) and whether they had any hearing and learning disabilities. The survey form is given in Appendix C.

After filling out the survey, participants were seated in front of the computer displaying the word *ready* on the screen and the experimenter explained the lexical decision task to subjects. They were told that they would be listening through headphones to English words and nonwords coming in a random order and instructed to press 'yes' and 'no' accordingly, corresponding to left and right SHIFT keys on the keyboard[1]. The speed of the response was emphasized. Subjects were prompted to press the ENTER key when they were ready to start the experiment. During testing the screen was blank. The first twenty items were practice trials given in order to familiarize participants with the task. After the 20 trials item presentation stopped and the screen displayed *ready* again. Presentation of the experimental trials started after the participant pressed ENTER again.

Upon completion of the lexical decision task, participants took the spelling test. They were instructed to listen to the recorded list of words followed by a list of nonwords and write them on a sheet of paper as they listen.

3.1.3. PARTICIPANTS

Sixty eight participants were recruited. All of them were undergraduate students of Cornell University (ages 18-22). Based on their answers to the survey, 18 participants were excluded from data analysis for the following reasons: 3 people reported having learning disabilities, 2 people were not native English speakers, the other 13 were early bilinguals and/or were born into a family whose primary language of communication was not American English, even though those participants claimed

[1] The keys were labeled

that English was their primary language at the moment of testing. All subjects received extra credit for one of their psychology courses for participation in the experiment.

3.2. RESULTS

3.2.1. LEXICAL DECISION TASK: ACCURACY AND REACTION TIMES

Five words (*Welch, strafe, rusk, thou, stein*) and five nonwords (*bune, hirm, letch, wog, binch*) were excluded from analysis due to high error rates (> 50%). The remaining stimuli were still matched on all the characteristics controlled for (average values shown in Table 2 are the values calculated after excluding the 10 stimuli). 10% of the data were lost due to the failure of the software to record reaction times (see section 3.3 for more detail). RTs beyond 3SD from the global mean (> 2203 ms and < 49 ms) were excluded from analysis (1.7% of the data). Mean error rates and RTs for consistent and inconsistent words and nonwords are presented in Table 3.

Table 3. Mean error rates and reaction times and standard errors (SE) for correct responses to consistent (CON) and inconsistent (INC) stimuli and differences (DIF) in error rates and mean RTs between consistent and inconsistent stimuli, Study 1.

	WORDS			NONWORDS		
	CON	INC	DIF	CON	INC	DIF
ERROR RATES	7.7%	7.2%	.5%	9.5%	12.8%	3.4%
REACTION TIMES (SE), ms.	1037 (7.5)	1069 (7.5)	32	1119 (8.1)	1139 (8.3)	20

The size of orthographic consistency effect was also calculated for each participant as the difference between mean RT for consistent word and mean RT for inconsistent words (Table 4).

Table 4. Range, mean and standard error (SE) of the size of orthographic effect across subjects, Study 1.

RANGE	MEAN	SE
-112 – 241ms	32ms	8.6 ms

Planned comparisons were performed the same way as in lexical decision studies reviewed in section 2: effect of consistency on mean error rates and correct mean RTs was tested by subject (t_1) and by item (t_2). Effect of consistency was not significant for error rates for words, but was significant for nonwords in the by-subject analysis. Consistency was not significant for RTs in the by-item analysis either for words or nonwords, whereas in by-subject analysis it was significant for words and marginally significant for nonwords (Table 5).

Table 5. Effect of consistency on error rates and mean reaction times of correct responses, Study 1.

	ERROR RATES T-RATIO	SIGNIF PROB
WORDS		
t_1 (49)	.61	.55
t_2 (53)	.21	.83
NONWORDS		
t_1 (49)	2.97	.0005*
t_2 (53)	.89	.38
	REACTION TIMES T-RATIO	SIGNIF PROB
WORDS		
t_1(49)	4.09	.0002*
t_2 (53)	1.02	.31
NONWORDS		
t_1(49)	1.93	.06
t_2 (53)	.76	.45

3.2.2. SPELLING TEST

3.2.2.1. Word Spelling

The number of misspelled words was calculated for each participant. It ranged from 0 to 10 (33%), with a mean of 3 (10%.).

3.2.2.2. Nonword Spelling

The scoring of the nonword spellings was based on their phonologically plausibility, i.e. whether each phoneme in a nonword was assigned a possible spelling. For example, for nonword /frek/, such spellings as *frake, fraik* and *fraic* would be considered phonologically plausible, whereas such spelling as *fraig* would be considered phonologically implausible. Such a scoring criterion is commonly used in assessing sublexical spelling processes in population with reading and writing dysfunctions (see, e.g., Rapp, Epstein & Tainturier 2002, Valdois, et al., 2003). However, for a normal adult population this criterion yielded non-informative results. Because in English multiple mappings onto graphemes exist for each phoneme and because the scoring criterion does not take into account the context of the grapheme/phoneme correspondence[1], the vast majority of spellings (97%) had to be scored as correct. Thus, this way of scoring was not sensitive enough to account for the variability observed in the data.

In order to assess participants' knowledge of sublexical phonology-orthography correspondences more precisely, the scoring was changed to account for spellings conditioned by context. Thus, a spelling was only scored as correct if both the onset and rhyme were assigned a valid spelling. An onset/rhyme spelling was considered

[1] For example, *c* is considered a correct spelling for /k/, even if it is followed by letter *e*, although in such a context *c* is always pronounced as /s/ in English, as in *cell*.

valid if there were other words in English in which such onset/rhyme was spelled the same way (these estimates were based on Ziegler et al.'s (1997) database).

The number of incorrect nonword spellings per participant ranged from 8% to 31% with a mean of 16%. These results suggest that in the majority of cases spellers did take the context of spelling into account.

To test the hypothesis that spelling proficiency predicted the size of the orthographic effect in individual subjects (Table 4), a regression model was fit to the data with the size of orthographic effect for each participant as the response variable and the number of misspelled words nonwords as two predictors. Neither of the predictors proved to be significant (Table 6).

Table 6. Effect of word and nonword spelling accuracy on the size of orthographic effect , Study 1.

SOURCE	F RATIO	SIGNIF PROB
# misspelled words	0.5744	0.4595
# misspelled nonwords	0.0402	0.8436

However, the power analysis showed that for both predictors, the power was small (.11 and .08 respectively), which makes the null results difficult to interpret.

3.3. DISCUSSION

The results of the study partially replicated previous findings on the effect of orthographic consistency of stimuli on auditory lexical decision: participants were on average 32 ms. faster on lexical decisions to consistent words than to inconsistent words, but were not more accurate.

The lack of effect on accuracy might have been due to the fact that all the words, except for the five excluded, were very familiar to the participants.

In word RTs analysis, the orthographic consistency effect was significant only by subject, but not by item. There may have been two reasons why the effect failed to reach statistical significance in both analyses. One reason is that the effect itself might have been weaker in the present data than in previously reported data – the difference between the mean RT for consistent and mean RT for inconsistent words was indeed smaller than in any previously reported results (Table 1). The other reason is that the variance in the data might have masked the effect.

Part of the large variance making the effect in the by-item comparison insignificant may have been due to the between-subject variability: as Table 4 shows, there were individual differences in the size of the orthographic effect, such that for some subjects the effect was not present or even worked in the direction opposite to the predicted. It is not known, however, how this compares to the other studies, since the published portions of findings do not include information on individual subjects.

There are also other two possible explanations for the variance in the data that have to do with the problems in the design and administration of the experiment.

- Even though a frame sentence was used in recording in order to keep the intonation constant, the speaker did not produce all the sentences in the same way, which resulted in inconsistent intonation across items. In addition, several participants reported that the speaker seemed to have certain particularities about her articulation that made some items less comprehensible.
- The testing room was not sound proof, therefore there is a possibility that the occasional background noises might have distracted participants.

There was also a problem related to data collection which makes some of the results of the study difficult to interpret. As noted above, 10% of the data were lost due to the failure of the software to record reaction times. The reason for that is not perfectly clear, but there is a possibility that in cases when a subject realized s/he had made an error after pressing YES or NO and pressed the correct one right after, the software interpreted it as a response to the next item. If this is the case, then the error rate data have to be discarded. Moreover, the 10% data loss added to the removed incorrect RTs and outliers resulted in the exclusion of over 20% of the data. This is a considerable and unpredictable manipulation of data, since the subsets of consistent and inconsistent words retained in the analysis were not the same across subjects. Such manipulation makes all results regarding individual differences, including the lack of association between spelling skills and the size of orthographic effects, difficult to interpret.

One other result is worth a brief mention. Orthographic consistency was found to affect nonword error rates and its effect on nonword RTs was marginally significant. This effect was only observed in one previous study with adults (Ventura et al., 2007) and seems to be less robust than orthographic effects on word RTs. In order to make sure that this was not an artifact of the present data set, this result has to be replicated in the same language with the same population before any conclusions about nonwords can be made.

To summarize, the study partially replicated orthographic effects on auditory speech perception previously observed in auditory lexical decision studies done in French, Portuguese and British English. However, the failure to fully replicate previous findings can not be easily interpreted because of problems related to the design, administration and data collection.

In order to address these problems, Study 2 was conducted, with a few changes made to the stimulus material, equipment and procedures.

4. STUDY 2

4.1. METHOD

4.1.1. STIMULI AND DESIGN

4.1.1.1. Lexical Decision Task

Stimulus list from Study 1 was modified in the following ways:

- Words that got high error rates in Study 1 were removed from the list.
- The most frequent words were excluded and more low-frequent words were added to reduce average frequency of words. This was done based on a speculation by Ziegler and Ferrand (1998) that the effect of orthographic consistency should be stronger for low-frequent words (see their discussion on p. 686). If it had been the case that the failure to fully replicate the results obtained in other languages was due to a weaker effect in English, using more low-frequent stimuli would make the effect more detectable.
- High error rates for five words in Study 1 suggested that those words were not familiar to the majority of the participants. Because familiarity and frequency, although not perfectly correlated, are not independent from each other, it was important to ensure that including more low-frequent words would not lead to the drop in average familiarity and to even more words with very high error rates. Therefore familiarity was included in the design as another control variable. A post hoc analysis of familiarity for stimuli in Study 1 revealed that all the words that had error rates higher than 50% had familiarity ratings lower than 4 on a 7-point scale, based on Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984, available online through Speech & Hearing Lab Neighborhood Database (http://128.252.27.56/Neighborhood/Home.asp). Therefore only words with familiarity scores higher than 5 were included.

• Nonwords that had high error rates were either excluded or modified to be less acoustically similar to real words. For example, nonwords that only differed from an existing word by voicing of one consonant (e.g., *burge – purge*) were changed.

The final list of items consisted of 50 words and 50 nonwords. Characteristics of stimuli are summarized in Table 7.

Table 7. Stimuli characteristics, Study 2.

	CONSISTENT	INCONSISTENT	SIGNIF PROB
WORDS			
# friends	2.9	2.8	.93
freq of friends	68.0	89.2	.29
# enemies	0	14.4	
freq of enemies	0	1781.0	
type consistency	1	0.08	
token consistency	1	0.17	
freq KF	7	7.1	.99
freq HAL	2835.1	2289.8	.29
freq CELEX	7.8	5.4	.22
familiarity	6.8	6.76	.61
# of letters	4.8	4.7	.73
#of phonemes	3.7	3.8	.57
duration (ms.)	531	480	.045*
# of orth neighb	5.96	4	.18
# of phon neighb	11.4	16.7	.07
# init phone freq	.06	.08	.19
# init biphone freq	.004	.006	.19
uniqueness point	4.2	3.9	.22
onset orth consistency	.91	.97	.25
onset phon consistency	.95	.96	.64
NONWORDS			
duration (ms)	511	497	.51
rhyme frequency (type)	199.0	214.0	.76
rhyme frequency (token)	8.0	7.5	.74

Recording and Programming

Stimuli were recorded at the same facilities and with the same equipment as in Study 1. A different speaker – a woman from Maryland – was invited to make the recording. The sentence frame was not used, but the speaker was instructed to read the list of stimuli trying to keep the intonation as constant and as monotonous as possible. Stimuli were recorded twice, which allowed to choose the most clearly pronounced samples and to further even out minor differences in fundamental frequency between items. Stimuli were normalized for peak amplitude in *Audacity 1.2.6* and edited in *Praat 4.6.21* to remove silence before and after each stimulus. Two native English listeners were asked to judge whether the finalized stimuli sounded natural and whether there were any particularities in the speaker's accent that impacted intelligibility. Neither of the listeners noticed anything unusual.

The presentation of stimuli and data collection were programmed in *E-Prime 1.0* on an Intel Pentium PC computer (for further details on hardware, see Appendix B). As in Study 1, the software was programmed to present stimuli with a 2 second interval and record the accuracy and reaction time of responses from the onset of auditory stimulus presentation to the moment when the subject pressed a response button. Prior to data collection, the experimental machine was tested by the *RefreshClockTest* developed by creators of *E-Prime* to check timing accuracy of experimental computers (Schneider, Eschman, Zuccolotto, 2002).

4.1.1.2 Spelling Test

In study 1, for a nonword spelling to be scored as correct, the onset and rhyme had to be assigned a possible spelling. However, such scoring criterion did not differentiate between the default (most frequent) spelling and other possible rhyme spellings. To make such differentiation possible and thus to assess spellers' sensitivity to probabilistic patterns in orthography more precisely, a new nonword list was created.

Another aspect of spelling proficiency that the new list of nonwords was meant to help assess was the sensitivity to phonology-orthography correspondences at different levels of granularity. Previous research (Juul, 2005; Perry & Ziegler 2004, Treiman, Kessler, & Bick, 2002), as well as the results of Study 1, have shown that when spelling nonwords, people demonstrate knowledge of phonology-orthography correspondences at both phoneme level and the level of larger subsyllabic units. However, even when the plausibility and frequency of a certain spelling are taken into account, it is not always possible to make conclusions about the size of units the speller was paying attention to. Consider, for example, one of the Study 1 nonwords /trok/. Spelling *troke* would be scored as both plausible and the most frequent spelling of rhyme /ok/. But at the same time, *o_e* is the default spelling of vowel /o/ (Kessler & Treiman, 2001). Therefore it is impossible to tell whether the speller used the most frequent phoneme-grapheme correspondence or took the larger rhyme context into account and thus to assess the aspect of spelling competence related to context sensitivity. In order to disambiguate these two cases, all the nonwords in the new list were constructed in such a manner that the default (or the most frequent) spelling of the rhyme would conflict with the default spelling of the vowel. For example, the most frequent spelling for vowel /e/ across all words containing it is *a_e*, as in *tape*. However, rhymes /el/, /en/ or /eθ/, as in *rail*, *pain,* or *faith*, are most frequently spelled with *ai* (Kessler & Treiman, 2002). Therefore, for nonword */flen/*, the default spelling of the rhyme would yield *flain*, whereas the default spelling of the vowel would yield *flane*.

Thus, the total number of each of the two kinds of the spellings in an individual's test would indicate how sensitive the speller is to phoneme-grapheme correspondences and to correspondences on the larger unit level.

All the data on default spellings of vowels and rhymes were taken from Kessler and Treiman (2001). The complete list of 25 nonwords is given in Appendix A. The list of irregularly spelled real words was the same as in Study 1.

A male speaker from California recorded test items. The recording was edited the same way as in Study 1.

4.1.2. PROCEDURE

- The procedure was similar to that of Study 1, except for the following modifications:
- A sound proof booth was used to test participants to minimize possible noise distractions. Unlike in Study 1, the experimenter was not present in the booth during testing.
- In addition to spoken oral instructions, participants also got written instructions on the computer screen. During stimuli presentation the screen was blank.
- Instead of using different hands to press 'yes' and 'no', participants used their preferred hand for both responses. Right-handed people used buttons '1' and '2' located on the numeric keypad of the keyboard, whereas left-handed people used the typewriter keys '1' and '2'.
- During spelling test, participants typed words into a text file using *MicroSoft Notepad* application instead of spelling words on paper. There were two reasons for this modification:
    - Several participants in Study 1 who gave their feedback said that typing for them had become a more natural activity than writing by hand.
    - Having participants spell by typing makes preparing data for analysis easier and reduces the probability of mistakes associated with entering the hand-written data into a computer file.

4.1.3. PARTICIPANTS

Fifty undergraduate students were recruited from Cornell University (ages 18-22). Based on the survey answers, 10 participants were excluded from data analysis: for 3 participants English was not the first or the only language that they acquired from their parents, the other 7 people acquired a second language early from their family members. Thus, only data from those subjects who were native speakers of English and learned a second language (if any) in a school environment were used. No participants reported any learning disabilities or hearing problems. Subjects either received extra credit for a psychology course or were paid $5 for participation.

4.2. RESULTS

4.2.1. LEXICAL DECISION TASK: ACCURACY AND REACTION TIMES

One word (*squad*) was excluded from the analysis due to error rate higher than 50%. The remaining words were still matched for all variables controlled for, except duration. Table 7 lists the summary characteristics for the two groups of stimuli calculated with *squad* excluded. RTs were trimmed according to the 3SD beyond the global mean criterion (> 1883 ms and < 56 ms), which accounted for 1.8% of the data.

Mean error rates and RTs for consistent and inconsistent words and nonwords are presented in Table 8.

Table 8. Error rates and mean reaction times and standard errors (SE) for correct responses to consistent (CON) and inconsistent (INC) stimuli and differences (DIF) in error rates and mean RTs between consistent and inconsistent stimuli, Study 2.

	WORDS			NONWORDS		
	CON	INC	DIF	CON	INC	DIF
ERROR RATES	6.7%	10.9%	4.2%	9.2%	7.8%	-2.6%
REACTION TIMES (SE), ms	915 (6.6)	929 (6.8)	14	964 (6.7)	965 (6.7)	-1

Planned comparisons were performed the same way as in Study 1 (Table 9).

Table 9. Effect of consistency on error rates and mean reaction times of correct responses, Study 2

	ERROR RATES T-RATIO	SIGNIF PROB
WORDS		
t_1 (39)	3.96*	.0003
t_2 (47)	1.81	.08
NONWORDS		
t_1 (39)	.72	.47
t_2 (47)	.46	.65
	REACTION TIMES T-RATIO	SIGNIF PROB
WORDS		
t_1 (39)	1.09	.28
t_2 (47)	.76	.44
NONWORDS		
t_1 (39)	.08	.93
t_2 (47)	.20	.84

The effect of consistency on error rates was significant for words in the by-subject analysis, but was not significant for nonwords. None of the comparisons revealed significant differences in RTs between consistent and inconsistent items. However, because for words duration was a confound potentially masking effect of consistency (consistent words were on average longer than inconsistent), these null results are difficult to interpret.

4.2.2. REACTION TIMES ADJUSTED FOR DURATION

To account for the confound, duration was treated as a covariate in the by-item analysis of RTs for words. With duration controlled for, the difference in RTs between consistent and inconsistent words was marginally significant. Mean RTs adjusted for duration and the ANOVA results are presented in Table 10 and 11 respectively.

Table 10. Adjusted mean reaction times and standard errors (SE) for correct responses to consistent (CON) and inconsistent (INC) words and the difference (DIF) in adjusted mean RTs between consistent and inconsistent words, Study 2.

CON	INC	DIF
902 (6.5)ms	941 (6.8)ms	39 ms

Table 11. Effect of consistency on adjusted correct mean RTs, Study 2.

	F-RATIO	SIGNIF PROB
F2 (1,46)	3.75	.058

Although by subject and by item analysis of mean RTs is a traditional way of analyzing data in psycholinguistic research, the by-item approach has been criticized in recent literature (Baayen, Davidson, Bates, in press; Baayen, Tweedie, & Schreuder, 2002) for not being very informative. The main problem with the by-item comparison, which is based on averaging over subjects, is that it does not account for the systematic variance due to subject-specific effects and thus treats this variance together with undifferentiated experimental error (Baayen et al., 2002). One of these effects is that some subject are on average slower than others. But more importantly, averaging over subjects does not account for the possibility that the factor studied in the experiment (orthographic consistency in the present case) does not affect responses of all subjects uniformly (i.e. that not all subjects take longer to recognize consistent words than inconsistent words). However, the analysis of orthographic effect sizes for individual subjects in Study 1 (Table 4) – and, as will be discussed later, in Study 2 as well (Table 16) – showed that this is exactly the case. Thus, because for some subject the effect of consistency works in the direction opposite to predicted, averaging over subjects masks the significance of the effect. The by-subject approach is more informative since it takes the subject-specific effects into account. However this approach does not account for any confound variables associated with individual items, which, in case such a confound exists, makes the results of the analysis hard to interpret.

In order to deal with these problems, a mixed effect model was fitted in the data with individual RTs entered as a response variable and consistency of items included as a fixed effect and subject and subject by consistency interaction included as random effects. The model is described by the following equation:

$$y_{ijk} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij} + \beta_2 W_{jk} + \varepsilon_{ijk}, \text{ where}$$

y_{ijk} – RT for subject *i*, word group *j*, item *k*,

β_0 – global average for consistent words,

b_{0i} – effect due to subject *i*,

β_1 – overall effect of orthographic inconsistency,

b_{1i} – effect of orthographic inconsistency specific to subject *i*,

X_{ij} – variable for word group: $X_{ij} = \begin{cases} 1 - inconosis\ tent \\ 0 - consistent \end{cases}$

β_2 – effect associated with word duration,

W_{jk} – duration of word *k*,

ε_{ijk} – error term.

Thus, like a by-subject t-test or ANOVA , this model accounts for the variance due to subject specific effects. The b_{oj} term accounts for some subjects being overall slower than others. Including the subject by consistency interaction term (b_{1i}*X_{ij}) into the model allows to test the hypothesis that effect of orthographic consistency is not the same for all. Finally, unlike a by-subject t-test or ANOVA, this model also allows to account for the duration covariate (W_{jk}).

The analysis was performed in *JMP 7.0*. statistical package. The relevant portion of the output is given in Table 12.

Table 12. Effects of consistency, duration, subject and subject by consistency interaction on individual correct reaction times, Study 2.

SOURCE	F RATIO	SIGNIF PROB
Consistency	11.9771	0.0012*
Duration	119.7967	<.0001*
Subject&Random	7.6620	<.0001*
Consistency*Subject&Random	1.7637	0.0027*

These results show that when duration of word and subject-specific effects are accounted for, the effect of orthographic consistency on reaction times is highly significant. Subject-specific effects tested by this model will be discussed in the next section.

4.2.3. INDIVIDUAL DIFFERENCES IN THE SIZE OF ORTHOGRAPHIC EFFECTS

The results of analysis of variance components associated with subject-specific random effect presented in Table 12 indicate that a) some subjects were significantly faster than others and b) that there were significant individual differences in how orthographic information affected reaction times to words.

The above analysis was performed on RTs of correct responses, which accounted only for 91.2% of the RTs for words. Therefore the consideration about unpredictable data manipulations discussed in section 3.3. still holds for this analysis.

To address this problem, the incorrect RTs were added to the dataset, however because the difference between error rates for consistent and inconsistent words was significant, accuracy was included in the model as a categorical predictor:

$$y_{ijk} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij} + \beta_2 W_{jk} + \beta_3 Z_{ijk} + \varepsilon_{ijk}, \text{ where}$$

β_3 – effect associated with accuracy

Z_{ijk} – dummy variable $Z_{ijk} = \begin{cases} 1 - incorrect \\ 0 - correct \end{cases}$

The results of the analysis are presented in Table 13.

Table 13. Effects of accuracy, consistency, duration, subject and subject by consistency interaction on reaction times, Study 2.

SOURCE	F-RATIO	SIGNIF PROB
Accuracy	11.1274	0.0009*
Consistency	13.4480	0.0007*
Duration	116.8523	<.0001*
Subject&Random	9.4314	<.0001*
Subject*Consistency&Random	1.6125	0.0099*

The subject-specific effects were significant with incorrect RTs included. The analysis also confirmed the significance of orthographic consistency on RTs.

As mentioned above, all the RTs beyond 3SD of the global mean were excluded. However the trimming affected only .8% of RTs to words, which makes data manipulation associated with excluded points negligible.

4.2.4. SPELLING TEST

Word spelling

Scoring for word spellings was the same as in Study 1. The number of misspelled words ranged from 0 to 10 (30%), with a mean of 2.8 (9%).

Nonword Spelling

For each subject, 4 parameters were calculated:

1. The number of default rhyme spellings,
2. The number of default vowel spellings,
3. The number of other possible rhyme spellings,
4. Other: spelling of real words instead of nonword (e.g., /kif/ - *Keith*), spellings suggesting misinterpretation of phonemes in the nonword (/delθ/ - *delf*), or typos.

Table 14 illustrates the scoring for the spelling of three nonwords by three subjects.

Table 14. Example of scoring of nonword spellings, Study 2.

SUBJECT	1	2	3
NONWORD			
/twaɪt/	twight	tweit	twight
/drel/	drail	drale	dreil
/pru/	prue	prew	proo
TOTAL			
default rhyme	2	1	1
default vowel	0	2	1
possible rhyme	1	0	1
other	0	1	0

Table 15 shows means and ranges for each of the 4 parameters across subjects.

Table 15. Mean counts and ranges for 4 different types of nonword spellings per subject, Study 2.

default rhyme	6.0 (2-11)
default vowel	8.6 (2-14)
possible rhyme	2.8 (1-5)
other	7.2 (2-13)

4.2.5. ASSOCIATION BETWEEN SIZE OF ORTHOGRAPHIC EFFECT AND SPELLING SKILLS

To test the hypothesis that the size of orthographic effect in individual's speech perception may be associated with various aspects of the individual's spelling competence, a regression model was fit to the data with the size of orthographic effect for each participant as a response and 4 predictors: the number of misspelled words and the number of 'default rhyme', 'default vowel' and 'possible rhyme' nonword spellings. (Table 17).

The size of the orthographic effect for each subject was first calculated the same way as in Study 1. Mean RTs for consistent and inconsistent words for each subject were estimated with incorrect responses included (to avoid data manipulation discussed earlier) and with adjustments for accuracy of responses and duration of words[1]. Then, the size of orthographic effect for each subject was standardized by dividing by the subject's standard deviation. Standardized and non-standardized values

[1] These adjusted means are calculated automatically by the statistical software when a model is fitted in the data with RTs as response variable, consistency, accuracy and duration as fixed factors and subject-specific effects as random factors.

of the size of orthographic effect are presented in Table 16. The two extreme values of the orthographic effect size were excluded from the regression analysis.

Table 16. Range, mean and standard error (SE) of the size of orthographic effect across subjects, Study 2.

	Range	Mean	SE
Non-Standardized	-14 – 110	42	3.9
Standardized	-.03 - .42	.25	.02

Table 17. Effect of 4 different aspects of spelling competence on the size of orthographic effect.

SOURCE	ESTIMATE	T-RATIO	SIGNIF PROB
# misspelled words	-0.021051	-2.44	0.0204*
# default rhyme	0.0049661	0.57	0.5746
# default vowel	-0.018883	-2.19	0.0364*
# possible rhyme	0.011168	0.67	0.5105

The results of regression analysis presented in Table 17 indicate that the size of orthographic effect was inversely correlated with the number of spelling mistakes in words and the number of default vowel spellings of nonwords. The positive correlation between the number of default rhyme spellings and the size of orthographic effect did not reach significance, but the power of the test of this correlation was very low (.09). Because default rhyme spellings and default vowel spellings were mutually exclusive, one might expect this positive correlation to be significant in a larger sample size. These results imply that the effect of orthographic consistency of rhyme in speech processing may be mediated by the strength of individual's orthographic representations (as measured by the number of correct word spellings) and the sensitivity to phonology-to-orthography correspondences at the levels larger than phonemes (measured by the nonword spellings based on regularities at the rhyme level).

4.3. DISCUSSION

Changes to the stimuli and procedures made to address problems of Study 1 seem to have helped slightly reduce the variance in the data and to make orthographic effects more detectable. Study 2 replicated previous findings by showing that orthographic consistency affected both accuracy and reaction time of lexical decisions for words in American English. Orthography did not seem to affect either reaction time or accuracy of lexical decisions for nonwords. Null effect for nonwords is consistent with most previous published results and therefore will not be discussed further (for interpretation of null effect for nonwords see Ziegler and Ferrand, 1998, and Ventura et al., 2007).

Because the effect of orthographic consistency on error rates is observed less reliably than on RTS in auditory lexical decision studies, it seems that it might be more dependent on a specific set of items – most probably on their familiarity – than the effect of consistency on RTs. However, the data available from this and other studies do not allow to test this hypothesis directly.

A mixed effect model fitted into the data showed that subjects varied significantly in the size of the orthography effect on their performance. For some subjects the effect was not observed and for some it worked in the opposite to predicted direction.

Data obtained in this study suggest that these individual differences may be associated with variability in spelling skills: subjects who made more errors when spelling words and subjects who relied more on the regularities on the grapheme-phoneme level rather than on the rhyme level when spelling nonwords exhibited less sensitivity to the orthographic consistency of rhyme in the lexical decision task. This finding is consistent with results of developmental studies showing that the influence of orthography on speech perception is a gradient phenomenon, the size of which is conditioned by listeners' experience with printed language. Present results add to the existing findings by showing that the size of orthographic effects is not uniform not only in children, but also in adults and depends not only on listeners' reading level, but also on their spelling competence.

5. GENERAL DISCUSSION

One goal of the present study was to extend the findings of orthographic consistency effects in auditory speech perception previously obtained in French and Portuguese to the population of adult American English speakers. The results confirmed that effects of orthographic consistency can be seen in English – a language that differs in orthographic transparency from the two previously studied Romance languages: English is more consistent than French but less consistent than Portuguese in phonology-to-orthography mappings, but much less consistent than both French and Portuguese in orthography-to-phonology mappings. This result is convergent with the recent findings from British English by Ziegler et al. (in press) who manipulated both orthographic consistency and pronunciation consistency of words and found that the latter affected neither reaction times nor error rates of lexical decisions on spoken words, whereas the former affected RTs, just like it does in French and Portuguese.

Direct comparison of the size of orthographic effects between languages was not among the goals of this study. In order to make these comparisons, it is necessary to obtain information on the characteristics of individual stimuli used in previous studies (e.g., frequency, neighborhood density, duration, etc.) and on performance of individual subjects (i.e., their RTs and error rates) in other experiments. The superficial comparison of the results of the present study to those obtained in different languages in the same experimental paradigm shows that the size of orthographic consistency effect on speech perception in American English falls in the range of previously obtained values. However more thorough statistical analysis is needed that would take into account cross-experiment differences in stimuli characteristics.

In sum, the replication of the orthographic consistency effect in auditory speech perception in yet another language once again confirmed that it is a real psycholinguistic phenomenon rather than an experimental artifact. The cross-linguistic corroboration of the results allows to make predictions for other alphabetic languages, such that in any language where sound-to-spelling mappings can be inconsistent, the effect of such inconsistencies should be expected to manifest itself in auditory speech processing. However the question about how the particularities of the orthographic system of each language would affect the size of orthographic effects in speech perception remains open.

Another goal of the study was to test the hypothesis that there might be differences in how the effect of orthographic consistency of words manifests itself in individual adult listeners. The results seem to confirm this hypothesis: the size of orthographic effect in the present study was strong for some subjects and non-existent or very weak for others.

Methodologically speaking, findings from the present study demonstrate once again that statistical procedures based on averaging across subjects are not always justifiable: pooling together subjects who are sensitive to the studied effects and those for whom the effects are week may lead to the failure to detect statistical significance of the effect of interest. Furthermore, the fact that the subjects demonstrate various degrees of orthographic influence on speech perception suggests that conclusions drawn from averaged results and generalized across the studied population may not be fully valid. Now that the existence of orthographic effects in auditory perception has been established, it is necessary to start to explore the nature of individual differences in the size of these effects.

The fact that different people behave differently in an experiment, and in particular, exhibit different degrees of sensitivity to the studied effect, is, of course, not at all surprising and may be due to a whole number of reasons not controlled in the experiment (attention span, engagement in task, activities preceding the participation in the experiment are just a few examples). Therefore conclusions regarding the significance of the individual differences warrant a certain amount of caution, especially considering that the effect of interest itself was quite small relative to the variance in the data. However, the fact that there were significant correlations between orthographic effect size in individuals' speech perception and the individuals' performance in a spelling test, suggest that these individual differences are systematic rather than due to pure chance.

In the present study, spellers who made more mistakes in irregular word spelling and were less sensitive to the context in nonword spelling exhibited less influence of orthography in lexical decision task. This finding calls for an explanation of associations between spelling and the size of orthographic effect. One possible explanation is that in better spellers orthographic representations are stronger, i.e. orthographic information is more salient, it is more easily accessed, more readily activated and thus impact speech perception more than in poor spellers. Such explanation suggest a causal relationship in which spelling competence determines orthographic effects in speech perception.

An alternative explanation, which is based on a different direction of causality, is also conceivable. One possible mechanism of automatic activation of orthography is the strong links between orthography and phonology established in the course of literacy acquisition and leading to integration of orthography and phonology both in visual and auditory language processing. Weaker integration of phonological and orthographic information may be responsible for both less orthographic effects in speech perception and worse spelling proficiency.

At this point, both explanations are based on speculations; more research is needed to determine exact mechanisms underlying the associations between spelling competence and orthographic effects in speech perception. Whatever the explanation is, the finding of these associations has implications for development: their existence confirms that the effects of orthographic consistency in speech perception are a result of learning to read and spell rather than anything else. However the parallels between literacy acquisition and the development of orthographic effects in speech perception are yet to be explored.

The present study has a number of limitations, which it shares with all previous studies of the kind. One limitation is that only monosyllabic words were used in the experiment. It is an open question, however, whether the mechanisms of monosyllable and polysyllable word processing are exactly the same.

Another limitation is related to the ecological validity of the experimental design: deciding whether something is a real word or not is not something that listeners often do naturally, therefore generalized conclusions regarding mechanisms of speech perception in general warrant some caution.

By finding associations between spelling proficiency and the size of orthographic effects, the present study adds to the existing (scarce) knowledge about the nature of the influence of orthography on spoken language. However, a lot of questions still remain unanswered. Thus, it is not clear whether orthographic effects in speech perception are in any way beneficial to listeners. In other words, does activation of orthographic information help listeners to recognize words? Does it interfere with speech perception? Or is it a side effect of literacy acquisition that is neither beneficial not detrimental for speech perception in adults?

Very little is still known about brain mechanisms supporting activation of orthography in speech perception, the development of those mechanisms in children learning to read and write and the relationship of those mechanisms with the mechanisms supporting reading and writing. These questions can not be addressed in behavioral experiments and require implementation of neuroimaging techniques.

To conclude, the present study adds to the knowledge about interactions of orthographic and auditory information in spoken language, but at the same time, the present findings raise a new set of questions regarding the nature of these interactions and underscore the need for further research into the complex processes going on in a literate mind.

APPENDIX A

MATERIALS

LEXICAL DECISION STIMULI, STUDY 1

Consistent words: barn, bath, bolt, branch, coin, couch, crisp, cult, dish, fierce, forge, fringe, globe, grasp, grudge, hoof, lodge, pledge, prep, ridge, roach, rusk, scarce, scratch, scribe, silk, slang, strafe, tube, Welch.

Inconsistent words: cloak, crane, crawl, cruel, crumb, dealt, debt, foam, heap, maim, mule, pie, plea, putt, quart, queer, scheme, screech, scroll, soap, squad, stern, stir, swear, stein, thou, tomb, tongue, trait, zinc.

Consistent nonwords: barm, binch, blive, cull, darch, dask, dount, doy, fab, fap, fench, fug, geft, grench, grick, hesh, letch, loth, noil, noof, prunch, rint, spuck, stig, tob, tump, wog, wouse, wush, zub.

Inconsistent nonwords: beel, beese, bef, borp, bune, burge, cripe, duff, durse, flomp, footh, fud, gruse, halm, hirm, jote, keef, lerth, maph, mose, murb, narse, prail, pulf, ralt, runge, sert, swence, torm, trute,

LEXICAL DECISION STIMULI, STUDY 2

Consistent words: boil, couch, crisp, cube, dull, fierce, forge, globe, grasp, grudge, harp, jog, lodge, marsh, mesh, munch, pledge, roach, scarce, shrimp, silk, soy, starch, tint, torch.

Inconsistent words: brawl, cloak, crane, cruel, crumb, deem, fern, freight, heap, maim, mule, pie, plea, putt, quart, queer, screech, starch, stir, stroll, swear, tomb, trait, yearn, zinc.

Consistent nonwords: barm, brive, clibe, coth, dask, dount, fap, fench, fug, geft, grench, grick, kinch, natch, netch, noof, scab, spuck, stig, swinge, tob, tump, wouse, wush, zub.

Inconsistent nonwords: beel, beese, borp, cripe, curge, duff, durse, flomp, footh, fud, gruse, jote, keef, lail, lirth, mose, murb, narse, pulf, ralt, runge, sert, swense, torm, trute.

SPELLING TEST STUDY 1

Words: ache, diaper, beige, bruise, bury, censure, choir, cough, coupon, coyote, crepe, depot, flourish, forfeit, gourmet, khaki, larvae, mourn, neuron, niche, plaid, quartz, quiche, scissors, suave, suede, synonym, syrup, waltz, yacht.

Nonwords: asoof, aspock, birker, blarse, bluzz, broise, celky, cempic, cousty, debuge, drune, fleekish, frake, garmic, gemple, gerner, goiled, javesy, jouling, keefy, kypist, mauzer, percle, pidget, prack, prizz, protch, quink, refedge, soned, spaunch, spile, spooch, spose, stike, strage, troke, querm, vutched.

SPELLING TEST STUDY 2

Nonwords: chined, dalth, domb, drail, drile, flane, foun, fra, hom, keith, kreeled, kreth, krowl, mutt, noach, pathe, prait, prue, roath, shree, spild, thawn, trut, tufd, twight.

APPENDIX B

CHARACTERISTICS OF RECORDING AND TESTING EQUIPMENT

Recording
Recording device: Marantz PMD660 digital sound recorder
Microphone: Electro-Voice RE-20
Bit Rate: 705 kbps
Audio sample size: 16 bit
Channels: 1 (mono)
Audio sample rate: 44kHz
Audio Format: PCM

Experimental machine Study 1
Model: Toshiba Satellite A100
Intel Celeron M
Processor: 1.70GHz
RAM: 736 MB

Experimental Machine Study 2
Model: Dell Optiplex GX620
Intel Pentium
Processor: 3.79 GHz
RAM: 1.99 GB

Headphones Study 1
Sony MDR - 7506

Headphones Study 2
Plantronics DSP 500

APPENDIX C

SURVEY FORM

Background information

1. Do you have any documented **learning disabilities** ☐ Yes ☐ No

   If yes, please specify ____

2. Do you have any **hearing problems**? ☐ Yes ☐ No
3. Where were you **born**? ________________________________
4. Was English the **first language** you learned growing up? ☐ Yes ☐ No

   and that you speak with your family?

   If no, please comment ____

   ____

   ____

5. Can you **sustain a conversation** in any language ☐ Yes ☐ No

   other than English?

   If yes, what language is that? ________________________________

   Please describe briefly when and how you learned it how you currently use this language (e.g. 'I speak it with my family members', 'I speak it when I go abroad', etc.)

   ____

   ____

   ____

6. Was English the **first language** you learned to **read** in? ☐ Yes ☐ No

   If no, please comment ____

   ____

7. Can you **read in any language** other than English? ☐ Yes ☐ No

If yes, please answer questions a.-d.

a. What language is that? ______________________________________

b. At what age did you start learning it? ___________________________

c. How well do you read in this language?

☐ I can read without a dictionary ☐ I can read with a dictionary

☐ I can only understand very basic phrases

d. How much do you currently read in this language:

☐ On a regular basis ☐ From time to time ☐ Very rarely ☐ Not at all

8. When you were just learning to read, you teachers most probably used a combination of methods to teach you. Which **method of teaching** reading did they **primarily** use?

☐ PHONICS (they taught you to connect individual sounds to letters

☐ WHOLE WORD (they taught you to recognize words as a whole)

☐ I changed schools - some used one and some used the other.

☐ I don't remember

Comments: ____

____

____

____

REFERENCES

Baayen, R.H., Davidson, D. J., Bates, D. M. (in press). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*.

Baayen, H. R., Tweedie, F. J., & Schreuder, R. (2002). The Subjects as a Simple Random Effect Fallacy: Subject Variability and Morphological Family Effects in the Mental *Lexicon. Brain and Language, 81*, 55–65.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Balota, D.A., Cortese, M.J., Hutchison, K.A., Neely, J.H., Nelson, D., Simpson, G.B., & Treiman, R. (2002). The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. http://elexicon.wustl.edu, Washington University.

Bruck, M. (1992). Persistence of Dyslexics' Phonological Awareness Deficits. *Developmental Psychology, 28*, 874-86.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in Phonological Recoding in German- and English-Speaking Children. *Scientific Studies of Reading, 2*, 31-54.

Chereau, C., Gaskell, M.G., & Dumay, N. (2007). Reading spoken words: orthographic effects in auditory priming. *Cognition 102*, 341–360.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

Cutler, A., Treiman, R., & van Ooijen, B. Orthographik inkoncistensy ephekts in foneme detektion? *Proceedings of the Fifth International Conference on Spoken Language Processing*, 2783-2786.

Juul, H. (2005). Knowledge of context sensitive spellings as a component of spelling competence: Evidence from Danish. *Applied Psycholinguistics, 26*, 249-265.

Kessler, B. & Treiman, R. (2001). Relationships between Sounds and Letters in English Monosyllables. *Journal of Memory and Language, 44*, 592-617.

Kreiner, D. S.,& Gough, P. B. (1990).Two ideas about spelling: Rules and word-specific memory. *Journal of Memory and Language, 29*, 103–118.

Kučera, H., & Francis, W.N. (1967). *Computational Analysis of Present-day American English*. Brown University Press.

Lindgren, S. D., De Renzi, E., & Richman, L. (1985). Cross-National Comparisons of Developmental Dyslexia in Italy and the United States. *Child Development, 56*, 1404-1417.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28,* 203-208.

Millier, K. M. & Swick, D. (2003). Orthography influences the perception of speech in alexic patients. *Journal of Cognitive Neuroscience, 15*, 981-990.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words* (Research on Speech Perception Progress Report No. 10). Bloomington: Indiana University, Psychology Department, Speech Research Laboratory.

Pattamadilok, C., Morais, J., Ventura, P., Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes, 22*, 700-726.

Perre, L., & Ziegler, J.C. (2008). On-line activation of orthography in spoken word recognition. *Brain Research, 1188*, 123-138.

Perry, C. & Ziegler, J.C. (2004). Beyond the two-strategy model of skilled spelling: Effects of consistency, grain size, and orthographic redundancy. *The Quarterly Journal of Experimental Psychology Section A, 57*, 325-356.

Rapp, B., Epstein, C., & Tainturier, M.-J. (2002), The integration of information across lexical and sublexical processes in spelling. *Cognitive Neuropsychology, 19*, 1-29.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's guide*. Psychology Software Tools, Inc.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143-174.

Shankweiler, D., Lundquist, E., Dreyer, L.G., & Dickinson, C.C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing*, 8, 267-294.

Slowiaczek, L. M., Soltano, E. G., Wieting, S. J., & Bishop, K. L. (2003). An investigation of phonology and orthography in spoken-word recognition. *The*

Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 56A, 233-262

Speech & Hearing Lab Neighborhood Database http://128.252.27.56/Neighborhood/Home.asp .

Taft, M., Castles, A., Davis, C., Lazendic, G., & Ngyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58*, 366-379.

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language, 47*, 448-468.

Tyler, M. D., & Burnham, D. K. (2006). Orthographic influences on phoneme deletion response times. *The Quarterly Journal of Experimental Psychology, 59*, 2010-2031.

Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing, 16*, 541-572.

Ventura, P., Morais, J., & Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: From sublexical to lexical involvement. *Cognition, 105*, 547-576.

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, *19*, 57-95.

Vitevitch, M.S. & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers, 36*, 481-487.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory & Language*, *47*, 1-29.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*, 683-689.

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition*, 32, 732-741.

Ziegler, J.C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental Science, 9*, 429–453.

Ziegler, J.C., Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: a developmental study. *Psychonomic Bulletin Review, 14*, 75-80.

Ziegler, J.C., Muneaux, M., & Grainger, J., 2003. Neighborhood effects in auditory word recognition: phonological competition and orthographic facilitation. *Journal of Memory and Language, 48*, 779–793.

Ziegler, J. C., Petrova, A., & Ferrand., L. (in press). Feedback consistency effects in visual and auditory word recognition: where do we stand after more than a decade? *Journal of Experimental Psychology: Learning, Memory & Cognition.*

Ziegler J.C., Stone, G.O., & Jacobs, A.M. (1997). What is the pronunciation for-ough and the spelling for/u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments & Computers, 29*, 600-618.